

小朱唇による構文解析器

林 寛子

1. 課題

次第立て言語「小朱唇」を使用して構文解析器の作成を試みた。カタカナ表記による単語分かち書きの文を入力し、その入力文に対し一連の処理を行い文法構造を解析し出力するシステムの設計である。文法は「文献3」で扱われているものを使用する。設計にあたっては構文の紛れをなるべく出すように努めたが不十分な点も多く、レベルとしてはオモチャ版程度である。

この構文解析器は、語詞解析用と構造解析用の二つの次第書きから成る。一つで済ませたかったが使用許可領域の関係で分割することになった。また、本稿で扱う次第書きが卒業論文提出後の改修新版であることを断っておく。改修の狙いは処理速度の向上と構造の簡素化であった。後者は口述試験時の指摘に基づき、前者は領域の幾分かの拡張を後日許されたため手を加えたものである。しかし、アルゴリズム自体に本質的な変更はしていない。

2. 使用文法について

今回計算機に与えた文法は「文献3」から取り出した文法規則に縮小・簡潔化や制限を強める等の手を加えたものである。取り出す際に、枠組み・基本となる規則を中心になるべく満遍なく取ることを念頭においた。使用した文法の文法番号を下表にする。表中の番号は「文献3」中の文法番号である。

表1 文献3から取り出した文法

構文名	文 法 番 号	構文名	文 法 番 号
文	B0	用連語	Y1
喚態句	K0, K1	相連語	S0, S1
述態句	Z0, Z1, Z2, Z3, Z10, Z11, Z12, Z13, Z14, Z15, Z20, Z25, Z26	情況語	X0, X1, X6, X10
		体連語	T0, T1, T2, T3, T10

(*) 実際の構文名はもっと細かく分けられている。上表では便宜上大きくまとめている。

計算機に与えた文法の形は本稿では省略する。

3. 解析作業の大筋

解析作業の大筋を下図に示す。実際は二つの解析部が別の次第書きになっている等の差異はあるが、設計思想の根本は次の通りである。

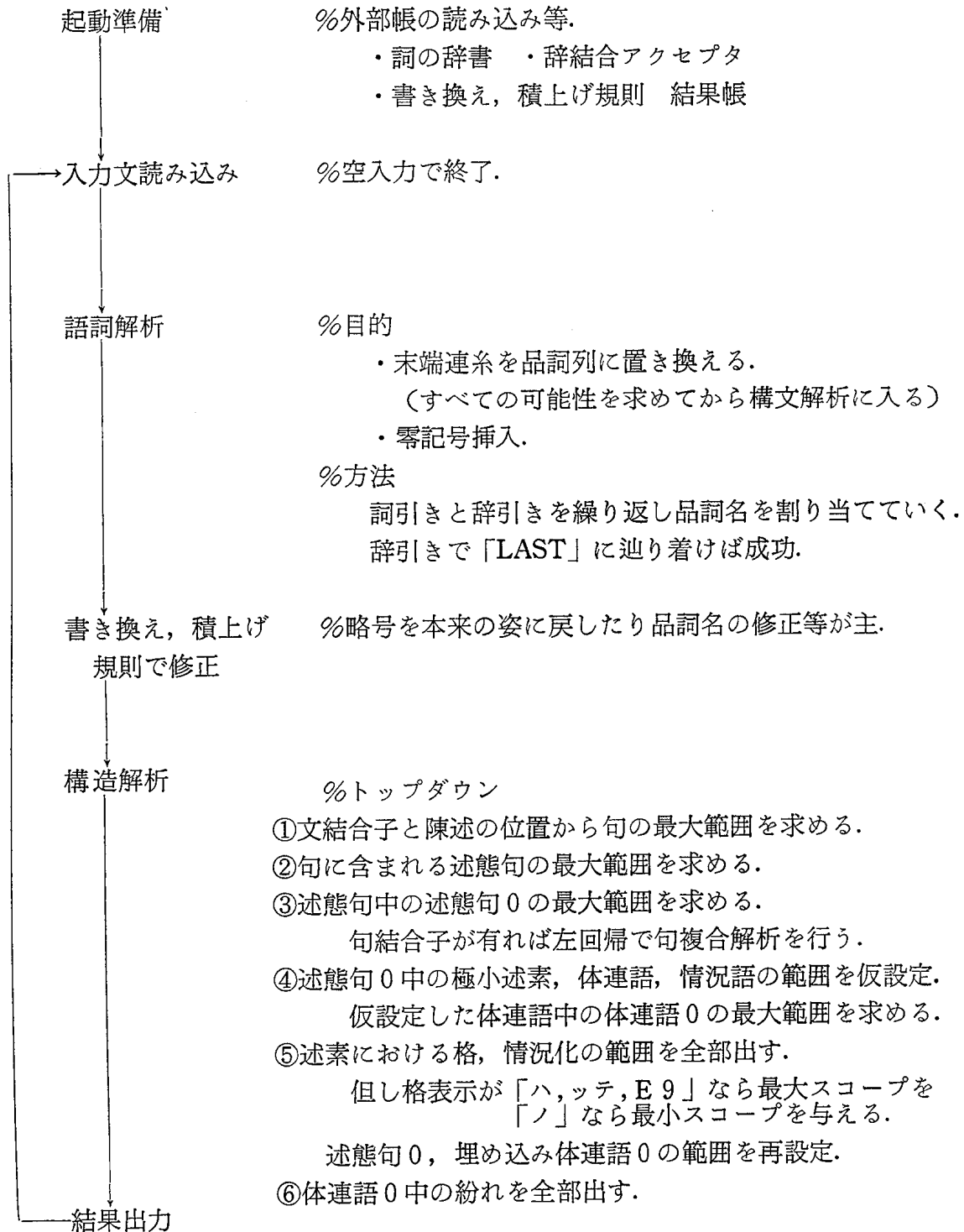


図1 解析作業の大筋

入力文読み込み

語詞解析

「タロウ ハ ソコ ニ イ 口 。

基本の動き

KAI='LAST' OR 'OUT' ナラ ヌケダセ。

LINEヲ#GO_「」_#LINEニ ワカテ。

(GO) 70SI. IIIKI.

KOTAEヲ#HINSI_「,..」_#CELLニ ワカテ。 %KOTAE=「(タイゲン@)..041」

HOZON_HINSIヲHOZONニ ウツセ。 %HOZON=「(タイゲン@)」

(CELL) 70 J I. HIKI.

```
%LINE='ハ ソコ ニ イロ .'
```

KOTAE7#KAI_「.」_#HINSIニ ワカテ.

```
%LINE='ソコニイロ...
```

HOZON_HINSIヲHOZONニ ウツセ。 %HOZON=「(タイゲン@) (カク) (カカリ@)」。

書き換え・積上げ

結果出力

カイセキ セイコウ! コスウ ハ 1

(タイゲン) (カカコ) (タイゲン) (カカコ) (Zソ) (Yレンゴ) (トウシ) (CZシ) (シヨVGE^2) (レイニシ) (テンカ1) \$

構造解析

図1の①～⑥のいずれの段階で処理されているのかの目安を示す。

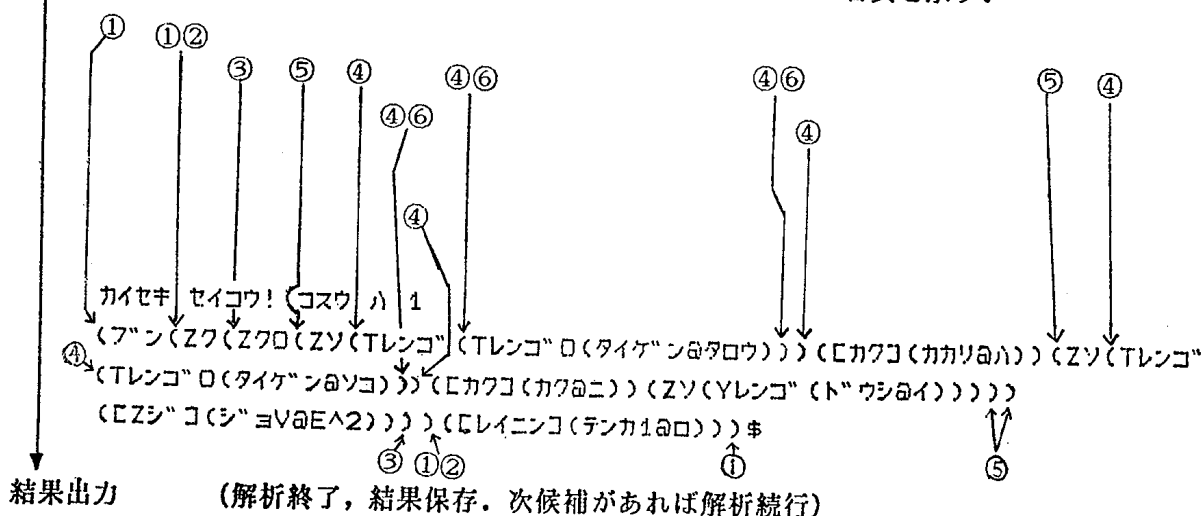


図2 一文処理の流れ

4. 辞書及び辞書様式一覧

本章では解析器がどのような品詞を認めているのか、また辞書様式がどのようなになっているのかについて説明する。

4. 1. はじめに

語は大きく詞と辞に分けることを前提とする。詞辞の境をどこに置くかは説の分かれるところであるが、本解析器では処理の便宜上文頭に立つことのできるものを詞とし、立てないものを辞とみなした。詞・辞を分け、詞の辞書は語幹辞書に、辞の辞書は辞結合アクセプタにもたせることにする。詞・辞を分けたのは両方で辞書の引き方を変えたためである。〔文献4〕で辞結合は3型オートマトンで表現できることが指摘されているが、今回の辞引きはその理論を取り入れた形で作成した。そのため、まず詞を語幹辞書で引き品詞名と共にアクセプタへの入りの番号 (cell番号) を受け取って次の連系からアクセプタを使って辞書引きを進めていく。アクセプタにはあらかじめ正しい結合を全部規定してあるのでかなり制限の強いものになっている。そのため、誤解析の発見が早く無用の検索が減り、また無意味な紛れを減らすことができるという利点がある。

4. 2. 辞書一覧

4. 2. 1. 詞について

下表の品詞を詞と認める。(品詞名は〔文献3〕のものによる。) 詞の辞書は可変帳でありユーザーによる語の登録・削除が可能である。

表2 詞の一覧表及び該当 cell 番号

品詞名	例	cell	品詞名	例	cell
文結合子	ガ, シカシ, ソシテ	011	連体詞	(*4)	013
感応言	アア, エエ, ウン	012	連体詞 0	ワガ, オノガ, コノ,	013
相体言	(*1)	042		ソノ, アノ, ドノ,	
形容詞	(*2)			コンナ, ソンナ, アン	
動詞	(*3)			ナ, ドンナ	
介テ複合動詞	ヤッテクル, モッテク		体言	(*5)	041
	ル, トンデクル		(場合)	バアイ, トキ, アイダ	041
動詞特類	イル, ミル, シマウ		(数量 0)	ダイブブン, イチブ	041
情態類 1	セツ, タン, オオイ	014	(数量 1)	ハンブン, フタリ	041
情態類 2	セツセツ, ドウドウ	015	(時分 0)	ヘイソ, ガンライ	041
情態副詞	マダ, スグ, ニコニコ	016	(時分 1)	イマ, キノウ, コトシ	041
程度副詞	ヤヤ, ズット, モット	017	形式名詞	ソウ, ヨウ	043

(＊1)いわゆる形容動詞語幹。(＊2)いわゆる形容詞。タイ類とゴトシは含めない。
 (＊3)いわゆる動詞。(＊4)いわゆる連体詞。(＊5)いわゆる名詞、代名詞、形式名詞のこと。

表2で空欄になっている活用語の cell 番号について規定する。活用語の場合、与える cell 番号は活用形によって変わってくる。また、詞の辞書は原則として語幹を見出し語とするが、カ変・サ変動詞は活用形を、形容詞「イイ」は終止形を、そのまま見出し語として登録する。対応は下表の通り。

表3 活用語用 cell 番号一覧

四段語尾	cell	備考	見出し語	cell	備考	一段語尾	cell	備考
ア段	1	(＊1)	コ	021	カ変	ル	033	
	2	(＊2)	キ	031	カ変	レ	027	
オ段	14		クル	033	カ変	φ	024	
エ段	028		クレ	027	カ変	形容詞語尾	cell	備考
ウ段	033	(＊3)	セ	022	サ変			
	0261	(＊4)	シ	023	サ変	カロ	14	
ン	025		スル	033	サ変	カッ	16	
イ／ッ	026		スレ	027	サ変	ク	029	
イ段	032	(＊5)	サ	5	サ変	イ	034	
シ	031		イイ	034	(＊6)	ケレ	027	
(＊1)「ナイ」の後続が可能なもの。 (＊2)「ナイ」の後続が不可能なもの。 (＊3)ウ音便を持たないもの。 (＊4)ウ音便を持つもの。(＊5)「イ、シ」は不可。(＊6)この形でしか使われないため。 (＊7)「よサそう、なサそう」等の「サ」。「美しい」等にはつかない。						(サ)	18	(＊7)
						φ	17	

4. 2. 2. 辞について

表4の品詞を辞と認める。(品詞名は「文献3」のものによる。) 辞は固定帳である辞結合アクセプタに載せてしまうので、あらかじめ、使いそうな語は網羅しておくことが必要である。ユーザーが使用する語は総て包含しているとの前提に立つので原則として語の追加・削除は認めていない。表4には今回載せた語を挙げておいた。無論語彙が全く足りないが、試作版であるのでこのくらいに絞った。また、今回のアクセプタには処理の都合上形式名詞の「ソウ」(伝聞)を「タ」に続く時に限り載せている。表中の品詞名に「＊」が付いているものは、「文献3」に品詞名の指定が無いので著者が独自に付け

た名称である。

表4 辞の一覧表

品詞名	内 容	品詞名	内 容	品詞名	内 容
不定辞	カ,カシラ	接続助詞	テ,デ,ガ,バ	格助詞	ガ,ノ,ヲ,ヘ,ニ
添加 0	ヨ,ヨネ,ヨナ,ネ,ナ,ゾ	モノノ類	モノノ		ト,デ,カラ,ヨリ
添加 1	ヨ,イ,イヨ,ロ,ロヨ	句接続詞	シカシ	係助詞	ハ
禁止	ナ,ナヨ	タイ類	タイ	副助詞	マデ
喚態モノノ類	モノ	*ガル	ガル	*ノ体連	ノ
助動詞(*)	ダ,アル,デス,マス,タ,ヨ	*テ	テ	*助詞	ッテ
	ウ,ウ,ラシイ,レル,セル,	*STカ	ソウ	*情況化	ニ,ト
	ラレル,サセル,ナイ ₁ ,	*ノ体連	ノ	*連体化	タル,ナル,ナ
	ナイ ₂ ,ベシ,ゴザル	*末カ	カ	*タバネ	ト

(*) [文献3] では「(ラ)レル,(サ)セル」を助動詞に含めていない。

4.3. 辞書様式一覧

語詞解析で使用する辞書には詞引き用と辞引き用の二種類ある。

前者は語幹辞書の形式を採っておりユーザーによる辞書項目の追加・削除が可能な可変帳になっている。詞引きが成功すれば品詞名と cell 番号の情報を返す。例えば、返された cell 番号が「1」なら、アクセプタの一行目に入ることになる。ところで初版では辞を総てアクセプタに載せており理論的にもその方が簡単であったが、結合リストが長くなりあまりにも作業能率が悪いので改修版ではかなりの部分を次第書きで手当てしている。番号の頭に「0」の付いているものがそれでこれらは直接アクセプタに入らず一旦次第書きでチェックを受けてアクセプタに入るものと入らないものとに振分けられている。

後者はリスト形式の固定帳になっておりユーザーによる項目の追加・削除は想定していない。辞の正しい結合経路をリストで示してあり合文法の結合なら適宜に品詞名あるいは零記号を受け取ってリストを脱出する。零記号の挿入はここで行う（一部句複合解析部でも行われている）。リストを脱出する指示には以下の三つがある。

・リスト脱出の指示

- ①「LAST」：解析成功、結果保存。次候補があれば解析続行。
- ②「OUT」： 解析不成功、結果破棄。次候補があれば解析続行。無ければ不成功の通報をし、次の文入力进行を待て。
- ③「KAI」： 辞引き一旦終了、結果保存。次連系から詞引きに入れ。

外部帳様式を図3に掲げる。

◎語幹辞書（詞引き用）

TEST.DIC :

ア*(ト"ウシ@),ラ.2,ロ.14,リ.032,ツ.026,レ.028,ル.033, * % ある

アア*(カンノウ@),.012, *

アノ*(レンタイ@),.013, *

イ*(ト"ウシ@),.024,レ.027,ル.033, (ト"ウシ@),7.1,オ.14,イ.032,ツ.026,エ.02 % 言う, いる

ウマレ*(ト"ウシ@),.024,レ.027,ル.033, * % 生まれる

オモ*(Sタイケン@),.042, *

オモイ*(タイケン@),.041, *

カ*(ト"ウシ@),7.1,オ.14,イ.032,ツ.026,エ.028,ウ.033, * % 買う

カケ*(ト"ウシ@),ラ.1,ロ.14,リ.032,ツ.026,レ.028,ル.033, * % 駆ける

カト*(タイケン@),.041, *

カフ*(タイケン@),.041, *

コ*(ト"ウシ@),.021, *

サク*(タイケン@),.041, *

ショホンハイ*(タイケン@),.041, *

ソコ*(タイケン@),.041, *

ソシテ*(フ"ンケツコ"ウ@),.011, *

ソラ*(タイケン@),.041, *

タカ*(ケイヨウシ@),カツ.16,ケレ.027,カロ.14,イ.034,7.029,.18, *

タハ"コ*(タイケン@),.041, *

タロウ*(タイケン@),.041, *

ニセンニンク"ライ*(タイケン"ニスウシ@),.041, *

↓ ↓ ↓

見出し語 品詞名 cell 番号

◎辞結合アクセプタ（辞引き用）

リスト中での品詞
は主に略号で表記
している。これは
助動詞の意。

NEWACCE : 200 行程

ナ"!80.ZVO, 2* % ナイの頭

ズ"!108.ZVO, 3* % セルの頭

ヌ"!116.ZVO, 4* % レルの頭

ネ"!101.ZVO, 5* % サルの頭

セ"!95.ZVO, 6* % ラルの頭

レ"!95.ZVO, OUT*

ヨ"!14 8*

ナ"!80.ZVO, 9*

サ"!5 10*

ラ"!6 11*

ズ"!108.ZVO, 12*

ヌ"!116.ZVO, OUT*

ネ"!101.ZVO, 9*

ウ"!123.ZVO, OUT*

タ"!66.ZVO, OUT*

タ"!66.ZVO, OUT*

カ"!44.(Cタイカコ@)(カ"ル@) 18*

ソウ"!KAI.(Cタイカコ@)(STカ@).043. OUT*

コ"!57.ZVO, OUT*

ア"!89.(シ"ヨV_アル@) 21*

ナ"!80.ZVO, OUT*

サ"!5 23*

ラ"!6 24*

ズ"!108.ZVO, 25*

ヌ"!116.ZVO, 26*

← 次の連系が「ナ」
なら80番目へ否なら
2 番目へ。

指示

見出し語 品詞名

図3 辞書の外部帳様式

5. 操作手順

最初に語詞解析用の次第書きを走らせる。まずタイトルと諸注意が画面に出る。次に外部帳の受け取りを始めるので、指示された外部帳名を端末から入力する。帳名の受け取りが済むと入力促進メッセージが表示されるので、端末から文を入力すればよい。文を入力すると語詞解析が始まる。解析が成功すれば、その個数と結果及び処理時間・実施日時・使用辞書名が出る。不成功の場合はその旨を通報する。一文の解析が終了すると再び文入力状態になるので実行したければ文を入力する。空入力で語詞解析を終了し結果帳の閲覧に移る。閲覧の希望はシステムが聞いてくるので答えればよい。以上で語詞解析は完了である。次に、構造解析用の次第書きを走らせる。前出の結果帳が今回の材料帳となる。構造解析部では材料帳が空になるまで処理を実行し、空になったら結果帳閲覧の希望を聞いて自動的に終了する。この解析結果が入力文に対する最終的な構文解析の結果である。但し、この出力のままでは大変見にくいのでこの結果帳を構文の木出力の次第書き（[文献1]所収）にかけるとよい。以上が操作の大筋である（後掲載の実行例参照）。下表にこれらの簡単なまとめを載せておく。

表5 操作仕様の概要

項 目	内 容
・用意する外部帳 （語詞解析）	詞の辞書，辞結合アクセプタ，書き換え・積上げ規則 結果帳（語尾表 改修版では不使用）
・入力文の仕様	次章参照
・語詞解析終了仕様	文入力状態で空入力
・語詞解析結果 の画面出力	成功の場合……解析結果の個数その結果及び処理時間 実施日時・使用辞書名 不成功の場合……不成功の旨を表示，他はほぼ同上
・用意する外部帳 （構造解析）	材料帳（語詞解析の結果帳がこれにあたる），結果帳
・構造解析終了仕様	材料帳が尽きれば自動的に終了
・構造解析結果 の画面出力	成功の場合……解析結果の個数その結果及び処理時間 実施日時（両解析の），使用辞書名 不成功の場合……不成功の旨を表示，他はほぼ同上
・結果帳の閲覧	語詞解析・構造解析共に可能。

6. 入力仕様

入力文は総て合文法であり未登録語はないものとの前提に立つ。入力仕様は表6の通

り。違反入力処理は処理過程で解析不成功になるであろうが、特にチェック機能を設けてはいない。

表6 入力仕様

項 目	内 容
入力方法	・端末から一文ずつキーイン。
入力文の表記	・単語分かち書き。文末記号には句点を使用する。
使用可能字種	・カタカナ、英数字、記号（但し「@」は不可）。
入力文の長さ	・理論上は無制限。実際は仕事領域の関係であり長くなっては無理。以下の例を目安に。 例：この裏町を覗く冷たいこぼれ陽よ。 この文の入力は可能であるが、 例：暗い浮世のこの裏町を覗く冷たいこぼれ陽よ。 では領域不足で事故になる。
読点許容箇所	・読点の使用は可能であるが、以下の箇所に限る。 ①格表示の直後 例：私、行きます。（零記号による表示） 例：私が、行きます。（格助詞による表示） 例：私からは、言えません。 （格・係助詞による表示） ②句結合の時、以下の三箇所 「開き」の直後 例：酒を買い、煙草も買う。 接続助詞の直後 例：行ったが、留守だった。 句接続の直前・直後 例：酒を買い、しかし、煙草は買わない。 例：行ったが、しかし、留守だった。

使用可能字種に漢字・平仮名が含まれていないのは、当時の計算機環境からくる制限に因る。英数字・記号を使用する際は無活用語の品詞による登録を想定しているが、語幹の字種のチェックをしていないので「NOWい」などの使い方も可能ではある。

「@」の使用を現時点では禁止しているが、理由はこれを次第書き中で制御記号として使用しているからであり「@」を使用できるように拡張することは容易である。

単語の単位は辞書への登録形が基準となる。従って「空飛ぶ円盤」を一単語とみなしたい時は辞書にこの語を体言で登録すればよいし、また三単語とみなしたいなら「空、飛ぶ、円盤」の三語に分けて登録することになる。

7. 実行例

下図4に実行例をしめす。「DIC」とは詞の辞書のことで、「COMB」とは辞結合アクセプタのことである。今積上げ規則は書き換え規則と同じファイルに入っている。

```

=== ショウジョウ タ"イ1.2ハ" [87-03-16] === (03/25/88 14 12 . 19)
** PARSE セ"ンハ"ン ** コ"シ カイセキ
チュウイ: フ"ン ハ タンゴ" ニ ワカチカ"キ セヨ。ズエ ニハ クラン ラ! オイル ニハ カイキ" ヨウ セヨ。
イマ シ"シヨ ヨウ ノ ハイル"ノ MAX ハ 70 だ" アル。

ク"カチヨウ キ"イ ハ ?
. W3
DIC キ"イ ハ ?
. D4
ヨ6
COMB キ"イ ハ ?
. COMB
ヨ6
カ#IRULE キ"イ ハ ?
. BUR
フ"ン ラ ト"ウヅ"!
. タロウ ハ ソコ ニ イ ロ。
シ"シヨニキ ニ ハイル
**シヨリ シ"カン 33.54" ヨウ **
カイセキ セイコウ! コズウ ハ 1
(タイケ"ン) ([カク] (カリ) ) (タイケ"ン) (Z)(Y) (L) (E) (N) (G) (ト"ウシ) ) ([Z] (シ" ヨV8EA2) ) ([L] (E) (N)
コ(チンカ16) ) #
セ"ンハ"ン 88キ" 03カ" ヲ 25ニキ カイシ シ"コク 14ジ" 147" 57" ヨウ DIC ハ D4
フ"ン ラ ト"ウヅ"!
.
セ"ンハ"ン ハ ココマチ" チ"ズ。 "PARSER コウ"ハ"ン" ニ ウ"ツ" チ クタ"サイ。

ココマチ" ノ ソウケツカ ラ 1:ミル ロ:イナ ?
. 1
*NEW*セ"ンハ"ン 88キ" 03カ" ヲ 25ニキ カイシ シ"コク 14ジ" 147" 57" ヨウ DIC ハ D4*7" シ*タロウ ハ ソコ ニ イ
ロ。
(タイケ"ン) ([カク] (カリ) ) (タイケ"ン) ([カク] (カリ) ) (Z)(Y) (L) (E) (N) (G) (ト"ウシ) ) ([Z] (シ" ヨV8EA2) ) ([L] (E) (N)
コ(チンカ16) ) #
オウリマズ。
===== Z 3 ===== (14 16 . 22)

```

図4の1 実行例

```

=== ショウシユシヨク 7" 41.2) 41.2) [87-03-16] === (03/25/88 14 17 . 5)
** PARSEER コウハシ ** コウ7" シ カイセキ
イマ ハイレツ / MAX ハ サ"イリヨウ カ" 150 ホカハ 60 テ"アル。

```

```

カキタ"シ FILE X4 ハ ?
. W2
サ"イリヨウ FILE X4 ハ ?
. W3
タイシヨウ ハ --> タロウ ハ ソコ ニ イロ。
コウホ ハ イマ 1
カイセキ
コウホ ハ イマ 1
ZS カイセキ
コウホ ハ イマ 1
TLEN"O カイセキ
TLEN"O カイセキ
**シヨリ シ"カン 28.58c" ヨウ **
カイセキ テイコウ! コズウ ハ 1
(7"シ(Z7(Z70(Z7(Y(TLEN"O(TLEN"O(タイゲ"ン(カカリヨウ))) (Cカ73(カカリヨウ))) (Z7(TLEN"O(TLEN"O(タイゲ"ン(カカリヨウ)))
(Cカ73(カ76ニ)) (Z7(Y(TLEN"O(ト"ウシヨウ))) (CZ7"コ(シ"ヨV&E^2))) (Cレイニ"コ(チンカ160))) #
セ"ンハ"ン 88ホシ 03カ"ツ 25ニチ カイシ シ"コク 143" 147"57c" ヨウ DIC ハ D4
コウハシ 88ホシ 03カ"ツ 25ニチ カイシ シ"コク 143" 177"25c" ヨウ

```

```

カイセキ ハ コレデ" ズ"チ オウリ テ"ズ。
ソウケツ"カ テ 1:ミル O:イサ ?
. 1
(7"シ(Z7(Z70(Z7(Y(TLEN"O(TLEN"O(タイゲ"ン(カカリヨウ))) (Cカ73(カカリヨウ))) (Z7(TLEN"O(TLEN"O(タイゲ"ン(カカリヨウ)))
(Cカ73(カ76ニ)) (Z7(Y(TLEN"O(ト"ウシヨウ))) (CZ7"コ(シ"ヨV&E^2))) (Cレイニ"コ(チンカ160))) #

```

```

オウリマス。
===== Z 3 ===== (14 19 . 4)

```

図 4 の 2

「 Tレンゴ〃 - Tレンゴ〃〇 - タイケン〃 == タロウ
 「 [カク] - カカリ == ハ
 ! 「 Tレンゴ〃 - Tレンゴ〃〇 - タイケン〃 == ソコ
 ! 「 [カク] - カク == ニ
 「 Zソ - Zソ - Zソ - Yレンゴ - トウシ == イ
 「 Zク - Zク〇 - [Zシ] - ショV == E^2
 ブン - [レイニン] - テンカ1 == 〇

図4の3

図4は三つの部分になっている。その二番目までがこの解析器での仕事。結果が括弧入れ表示ではわかりにくいので、[文献1]にある次第書きで木表示にした。

8. 評価

ここに述べた解析器の性能につき、以下の点には改良の余地がある。

1. 連体修飾の紛れの検出がなお不十分——連体修飾が格や述態辞を飛び越えて離れた体連語にかかる場合の処置が十分でない。例えば「あの将軍だった人」の解析で「あの」は「将軍」にしか掛けてない。紛れが爆発的にふえる懸念があるにせよ、それをどう抑えるかの問題を含めて、この扱いを講じていない。

2. アクセプタの正確性——辞結合をチェックするアクセプタについて、この考え方の理解が十分かどうか自信がない。これを活かすには、次第立ての技術より前の国語学的分析をもっとよく行う必要があると思われる。

3. 処理速度が遅い——小朱唇処理系があまり速い処理系でないにしても、次第立てを工夫すればこれほど遅くはならないであろう。

解析方式についても、今回のトップダウンのほかにボトムアップがあり、トップダウンがこれこれの点ですぐれているからという確信があって選んだ訳ではなかった。次第立ての過程ではボトムアップならもっとしやすかったのにとすることもあり、この辺のことは事前によく検討しておくべきだったと反省する。両方式のものを作って比較するだけの余裕がなかったのが残念である。

参考文献

- [1] 水谷静夫(1986)『次第立て言語小朱唇の手引』東女大日本文学科。
- [2] ——(1986)『小朱唇言語仕様』東女大日本文学科。
- [3] ——(1983)国文法素描、『朝倉日本語新講座3 文法と意味I』, 朝倉書店, pp. 1-80.
- [4] ——(1971)リスト処理による活用アクセプタ, 『計量国語学』[56], pp. 6-29.
- [5] ——(1974)再び文法と意味と, 『bit』6[6], pp. 424-430.

(はやし ひろこ 昭和63 日文卒)